Section 3.0 Statistical Methodology for Establishing Baseline Conditions and Setting Discharge Limits at Remining Sites

3.1 Objectives, Statistical Principles and Statistical Issues

The Rahall amendment (CWA Section 301(p)) states: "... *in no event shall such a permit allow the pH level of any discharge, and in no event shall such a permit allow the discharges of iron and manganese, to exceed the levels being discharged from the remined area before the coal remining operation begins.*"

"Levels" is interpreted here to mean the entire probability distribution of loadings, including the average, the median, and the extremes. If P percent of loadings are ≤ some number Lp during baseline, then no more than P percent should be ≤ Lp during and after remining. For example, if during the baseline period, 95 percent of iron loadings are ≤ 8.1 lbs/day and 50 percent are ≤ 0.3 lbs/day, then during and after remining the same relationships should hold true. This should hold true for pH, and for loadings of acidity, iron, and manganese.

The objective of Section 3.0 is to provide statistical procedures for deciding when the pollutant levels in a discharge exceed the levels at baseline. These procedures are intended to provide a good chance of detecting a substantial, continuing state of exceedance, while reducing the likelihood of a "false alarm." To do this, it is essential to have a sufficiently large number of samples during and after baseline. The methods provided here may be applied to either pH or pollutant loadings.

The procedures described below will provide limits for both single observations and annual averages. This is intended to provide checks on both the average and extreme values. There is a need to take into account the number of observations used to determine compliance when setting a limit or when otherwise determining compliance with baseline. For example, the collection of a greater number of samples from a discharge will reduce the variability of the average level (provided that samples are distributed randomly or regularly over the sampling year).

Use of a statistical decision procedure should result in suitable error rates. Technically these are usually referred to as the rate alpha (α) for Type I errors and the rate beta (β) for Type II errors. The error of concluding that an exceedance has occurred when the discharge is exactly matching the baseline condition is intended to happen with probability α no more than 0.05. When the discharge level is substantially less than baseline, the probability of making this error is expected to be very low. This probability could be controlled for each decision, or it could be controlled for *all* decisions for one analyte (e.g., 48 tests of a maximum daily limit and four tests of an annual average over a four-year period). When many decisions will be made, the overall error rate is a concern. The error of concluding that no exceedance has occurred when the discharge has in fact exceeded baseline levels, is intended to happen with probability β no more than 0.25. Again this could be controlled for each decision or over all decisions during the life of a permit.

There is significant, positive autocorrelation of flows, concentrations, and loadings in mine discharges over periods of 1-4 weeks (Griffiths, in preparation). Sample estimates of the variance, used in the statistical procedures proposed here, are inaccurate unless adjusted for autocorrelation. Without adjustment, variance is underestimated. Such adjustments are discussed by Loftis and Ward (1980), and EPA (1993) has accounted for autocorrelation in a previous effluent guideline. Such adjustments require an estimate of the autocorrelation coefficient. However, one cannot reliably estimate site-specific autocorrelation from small sample sizes (e.g., n=12). Thus, there is a need for default adjustment factors for sample variance. These factors would be developed from estimates of autocorrelation, using data from coal mine discharges having sufficient data obtained over the course of at least two years. Such factors may or may not be specific to all types of mine discharge. EPA has provided a default value of first-order autocorrelation $\rho_1 = 0.5$ to be used in calculating an adjustment. EPA may provide a different value in a final rulemaking and in the final version of this Statistical Support Document, after analysis of data and comments provided by states and other sources.

One year of sampling may not adequately characterize baseline pollutant levels, because discharge flows can vary among years in response to inter-year variations in rainfall and ground water flow. There is some risk that the particular year chosen to characterize baseline flows and loadings will be a year of atypically high or low flow or loadings. Permitting authorities should be aware of this risk and may want to inform permittees of this risk in order to encourage multi-year characterization of baseline. There is a need to evaluate the differences among baseline years in loadings and flows, based on further analysis of data now available to the Agency. Using such information, EPA may provide optional statistical procedures in a final rulemaking that could account for the uncertainty in characterizing baseline from one year, or that could account for the unrepresentative character of a baseline sampling year. Such procedures could employ modifications of the proposed statistical procedures that use estimates of the variance among baseline years in loadings, developed from long-term datasets, or could employ adjustments to the baseline sample statistics to account for a baseline sampling year that was atypical in rainfall or discharge flow. Such an adjustment could be a factor (multiplier) or a statistical equation estimated by regression.

The proposed statistical procedures are intended to provide environmental protection and to ensure compliance with the Rahall amendment. EPA has not yet evaluated the error rates of these decision procedures. EPA intends to evaluate the decision error rates of each procedure by computer simulations. Depending upon comments and associated evidence, and depending upon EPA’s further evaluations, EPA may modify or reject any of these procedures, or may change the recommended number of samples, in order to provide suitable error rates.

3.2 Statistical Procedures for Calculating Limits from Baseline Data

Two alternative statistical procedures are described, A and B. These may be modified to require accelerated monitoring (Procedure C) after a warning level or condition is exceeded.

3.2.1 Procedure A

Procedure A is a modification of the methodology used by the State of Pennsylvania. Computational details appear in Figure 3.2a. Pennsylvania monthly and annual average checks are defined as follows:

<u>Monthly (or single-observation maximum) check</u>: A tolerance interval is estimated for the baseline loadings (for $n < 17$, the smallest and largest observations define the interval endpoints). The baseline upper bound (usually the maximum baseline loading) is the value of interest. Two consecutive exceedances of the upper bound trigger weekly monitoring. Four consecutive exceedances during weekly monitoring trigger a treatment requirement. Thus, six exceedances must occur consecutively before a treatment requirement is triggered.

<u>Annual average check</u>: A robust, asymptotic estimator of an 95 percent confidence interval for the median is calculated for the baseline period and a post-baseline period; if the post-baseline interval exceeds the baseline interval, an exceedance is declared.

EPA strongly recommends that Pennsylvania's procedure be modified so that corrective actions are triggered after three exceedances of the single-observation maximum have occurred within a two-month period. Weekly monitoring should be initiated promptly, within 7-10 days, after a single exceedance. Occurrence of two more exceedances (whether consecutive or not) during this weekly monitoring should then result in appropriate and effective corrective actions. In a final rule and in the final version of this Statistical Support Document, EPA may evaluate other decision rules and may identify decision rules that provide a reasonable balance between the two decision error rates.

This modification would increase the ability to detect exceedance of the baseline pollutant level, while providing sufficient protection against unnecessary monitoring. Pennsylvania's single-observation check requires six consecutive exceedances to trigger treatment. Even when the probability of an exceedance during remining is high, say 0.5, the chance of seeing six consecutive exceedances is still low. For example, after one exceedance, the odds of seeing five more consecutive exceedances is about 1:31 (assuming no serial correlation; in practice the odds may well be higher). The modification provides substantially better protection. The risk of unwarranted monitoring is still low. When the observed maximum of twelve baseline loadings is used as the single-observation maximum limit, the probability is 0.95 that this value is at least as large as the true baseline 77.91-th percentile, and the probability is 0.90 that this value is at least as large as the true baseline 82.54-th percentile. Therefore, when the baseline pollutant level distribution is not being exceeded during remining, then with 95 percent confidence the probability is no more than 0.214 of observing two exceedances in four weekly observations, and with 90 percent confidence the probability is no more than 0.143 of observing two exceedances.

Comments: (a) The annual intervals are robust but are not non-parametric. A suitable non-parametric procedure would apply the Wilcoxon-Mann-Whitney test to compare baseline to post-baseline periods; use of this test would not greatly decrease, and could increase, statistical power. (b) The annual intervals depend upon an asymptotic approximation and the intervals are symmetric. Loadings data for pre-existing discharges are highly asymmetric, and annual means and medians are likely to be somewhat asymmetrically distributed. Therefore, suitability of the approximation needs to be evaluated for small samples (e.g., $n = 12$). (c) The single-observation check uses a non-parametric estimate of a percentile, equivalent to a 1-sided tolerance bound. For $n = 12$, and using the maximum baseline observation as the upper bound, the probability that a proportion P of the baseline distribution is not greater than this maximum is $1 - P^{12}$. For example, the probabilities are 0.93, 0.72, and 0.46 that 80 percent, 90 percent, and 95 percent, respectively, of the baseline distribution lies below the observed maximum. Therefore, twelve baseline observations provide a rather imprecise tolerance interval. (d) As previously stated, the single-observation check requires six exceedances before treatment is required. It might be better if one exceedance triggered prompt follow-up monitoring for confirmation, and one or two additional exceedances lead to a site investigation and remedial or corrective actions.

3.2.2 Procedure B

Procedure B consists of three checks: an upper limit on single observations, a yearly test of the mean or median, and a cumulative monthly evaluation with a cusum test. Computational details of Procedure B are provided in Figure 3.2b.

The single-observation limit is a parametric estimate of the 99th percentile of loadings, developed using baseline data. The method of calculating this limit assumes that loadings are approximately log-normally distributed (using natural logarithms). The methodology is similar to that used by EPA to calculate "maximum daily" limits for effluent discharges. It would be acceptable to substitute a non-parametric estimate of a high percentile if at least 50 baseline data results were available. That estimate would be the k-th largest of n data, and would estimate the 100k/(n+1) percentile; e.g., the largest of 50 observations is a non-parametric point estimate of the 98th percentile.

The annual test of the average or median employs either the t-test (using the natural logarithms of the loadings) or the non-parametric Wilcoxon-Mann-Whitney test.

The cumulative monthly evaluation employs a cusum test. This is expected to detect an increase in the mean loading somewhat sooner than the annual test in many cases. More important, it is expected to provide better detection of a long-term gradual increase and a long-lasting step increase than would be provided by the annual test. The cusum test as employed here, requires that data be approximately normally distributed; for that reason, transformation of loadings would be required. Computational details appear in Figure 3.2b. Procedure B also calls for reporting time-series plots of monthly data and the monthly cusum statistic, along with the baseline statistics.

Comments: Accuracy of the parametric estimates depends on approximate log-normality of the data. If the data are even more heavy-tailed (e.g., gamma distribution), this procedure could under-estimate the percentiles; if less heavy-tailed, it could over-estimate percentiles. The data are highly asymmetric, and annual means are likely to be somewhat asymmetrically distributed. Therefore, suitability of the approximation needs to be evaluated for small samples (e.g., n=12). Estimates of ρ_1 should not be made for individual discharges using n=12 data. Reliable estimates require a larger number of data, possibly from a different sample than that used to estimate the mean and variance ($n_\rho > 30$ is necessary). It should also be possible to use the average of ρ_1 values calculated for a number of discharges having similar flow and concentration relationships.

3.2.3 Use of Triggered or Accelerated Monitoring in Procedures A and B

Triggered or accelerated monitoring (Figure 3.2c) can be applied with Procedure A or B. This consists of accelerated monitoring after a single exceedance of baseline, or after an exceedance or continued exceedances of the cusum warning level. Triggered or accelerated monitoring provides a way to confirm an exceedance seen during routine monitoring. A decision is based on the new monitoring results. Accelerated monitoring would begin promptly (within 7-10 days of the exceedance) and would be conducted weekly or more frequently, for 3-6 sampling occasions. Accelerated monitoring (if used as a condition or option for determining non-compliance) could guard against a declaration of non-compliance on the basis of a transient exceedance, and would provide a means to demonstrate continuing exceedances. It could guard against the possibility of instituting expensive remedial measures when there was no continuing exceedance of baseline conditions, or when simpler remedial measures can quickly be implemented.

Figure 3.2a: Procedure A: both tests 1 and 2 are applied

x_i = pollutant loading measurement (product of flow and concentration measurements)
n = number of x_i results in the baseline dataset

1. Single-observation trigger

Order all n baseline measurements such that $x_{(1)}$ is the lowest value, and $x_{(n)}$ is the highest.

If n<17, then:
The single-observation trigger will equal the maximum baseline value, $x_{(n)}$.
If n>16 then:
Calculate the sample median (M) of the baseline events:
If n is odd, then M equals $x_{(n/2+1/2)}$.
If n is even, then M equals $0.5*(x_{(n/2)} + x_{(n/2+1)})$.
Calculate M_1 as the median between M and the maximum $x_{(n)}$.
Calculate M_2 as the median between M_1 and $x_{(n)}$.
Calculate M_3 as the median between M_2 and $x_{(n)}$.
Calculate M_4 as the median between M_3 and $x_{(n)}$.

The single-observation trigger L equals M_4.

If, during remining, an observation exceeds L, proceed immediately to weekly monitoring for four weeks (four weekly samples). If, during weekly monitoring, any two observations exceed L, declare exceedance of the baseline distribution.

2. Annual test

Calculate M and M_1 as described above.
Calculate M_{-1} as the median between the minimum $x_{(1)}$ and the sample median.
Calculate $R = (M_1 - M_{-1})$.

The subtle trigger (T) is calculated as:

$$M + \left[\frac{1.58 * [(1.25 * R)]}{[1.35 * \sqrt{n}]} \right]$$

Calculate M' and R' similarly for a year's data during re-mining. Calculate $T' = M' - (1.58*1.25*R')/(1.35\sqrt{n'})$. If $T' > T$, conclude that the median loading during re-mining has exceeded the median loading during the baseline period, and declare an exceedance.

Figure 3.2b: Procedure B: All three tests or limits are applied.

x_i = pollutant loading measurement (product of flow and concentration measurements)
y_i = $\ln(x_i)$
n = number of observations, y_i, in the baseline dataset
E_y = $\Sigma (y_i) / n$ $\quad 1 \le i \le n$
S_y^2 = $A \ \Sigma [(y_i - E_y)^2] / (n - 1)$ $\quad 1 \le i \le n$

A = $[1 - (2/n) r_1]^{-1}$
where r_1 is an estimate of the first-order autocorrelation (ρ_1) of y_i (not x_i).
(Procedure B will use $r_1 = 0.5$)

E_X = $\exp (E_y + 0.5 S_y^2)$

Z_{95} = 95th percentile of standard Normal distribution = 1.6449
Z_{99} = 99th percentile of standard Normal distribution = 2.3263

The estimates of mean and variance above are used when data have no "below-detect" observations [1,2]. When some observations are "below-detect," there are a number of approaches to estimation. If there is only one reported quantitation limit for the below-detect observations, one could apply the delta-lognormal procedures of the TSD [1,2](Appendix E). If there are multiple quantitation limits for the below-detect observations, (1) one could use as data either the reported quantitation limits (i.e., analytical minimum levels, MLs), (2) one-half their reported values, or (3) one could use Regression on Order Statistics (ROS) to estimate the mean and standard deviation of (natural) logarithms [3,4].

1. Single-observation limit = $\exp(E_y + Z_{99} S_y)$

Single-observation warning level = $\exp(E_y + Z_{95} S_y)$

Consider making an investigation and taking action when the warning level is reached.
Keep and report a chart showing x_t vs. month or successive observation number, and showing the warning level and the single-observation limit.

(Figure 3.2b continues)

[1] H.D. Kahn and M.B. Rubin, 1989, "*Use of statistical methods in industrial water pollution control regulations in the United States*," Environmental Assessment and Monitoring 12: 129-148

[2] "*Technical Support Document for Water Quality-based Toxics Control*," March 1991, EPA/505/2-90-001, Appendix E

[3] Helsel, D.R., and T.A. Cohn. 1988. Estimation of Descriptive Statistics for Multiply Censored Water Quality Data. Water Resources Research. Vol. 24, No. 12:1997-2004.

[4] Hirsch, R.M., and J.R. Stedinger. 1987. Plotting Positions for Historical Floods and Their Precision. Water Resources Research. Vol. 23, No. 4:715-727.

Figure 3.2b (continued): Procedure B

2. Cumulative Sum (Cusum) test (decision interval) [5]

Limit value: C_t must be less than H - declare an exceedance if C_t reaches or exceeds H.

Set $C_0 = 0$. Index each new datum sequentially (e.g., t = 1 to 60), and calculate

$C_{t+1} = MAX\{0, C_t + (Y_t - K)\}$,

that is, the new sum C_{t+1} is $C_t + (Y_t - K)$, if that is positive or else it is zero.

Calculate $K = E_y + f S_y$, using $f = 0.25$.

Calculate $H = h S_y$, using $h = 8.0$.

Cusum Warning level : Also calculate $W_t = MAX\{0, W_t + (Y_t - K_w)\}$, as done for C_t, using

$K_w = E_y + f S_y$, using $f = 0.5$.

$H_w = h S_y$, using $h = 3.5$.

A warning is indicated if W_t reaches or exceeds H_w.

Keep and report charts showing C_t and W_t vs. month or successive observation number, and showing the Cusum Limit and warning levels H and H_w. Consider making an investigation and taking action when the warning level is reached.

3. Annual comparison [4]

Compare baseline year loadings with current annual loadings using the Wilcoxon-Mann-Whitney test [6] for two independent samples. Alternatively, use the two-sample t-test with log-transformed data (using natural logarithms), y_t. Use a one-tailed test with alpha 0.025 to 0.05.

---

[5] Wetherill, BG & Brown, DW, 1991, Statistical Process Control, Sections 7.1.7 and 7.2.1, and Table 7.6.

[6] See Conover, W.J., 1980, Practical Nonparametric Statistics, 2nd ed., and other textbooks.

Figure 3.2c: Accelerated (Triggered) Monitoring.

The following could be implemented as a permit condition that requires weekly monitoring after one or more exceedances of a daily maximum level or a Cusum warning level during routine monitoring (monthly monitoring is assumed to be the routine). Accelerated monitoring is intended to confirm the continuation and the magnitude of a state of non-compliance, not to confirm the first exceedance (which is in fact an exceedance). Thus, accelerated monitoring is best triggered after a warning *level* is exceeded, and before a legally enforceable *limit* is exceeded. Professional judgement may be exercised in interpreting the data. For example, did the exceedance coincide with record flows and rainfall ? Do the accelerated monitoring data suggest a return to baseline levels, or a trend of rapid decrease ?

Accelerated (Triggered) Monitoring.

Promptly, within 5-10 days after an exceedance of a maximum level or a warning level (below), begin weekly monitoring. Require monitoring for three to six weeks in all; six observations are recommended to quantify the loading. Declare confirmation after observing one exceedance. The probabilities of false positives and false negatives can be inferred from Table 3.2a (which applies to uncorrelated data; we expect the probabilities to be higher for positively correlated data).

Application to Procedure A.

Single-Observation Maximum Level. Apply this procedure after the single-observation maximum limit is exceeded.

Application to Procedure B.

Single-Observation Maximum Level. Apply this procedure after the single-observation maximum limit is exceeded.

Cusum warning level. Apply this procedure after the warning level H_w in the Cusum test is exceeded once, or twice in succession. Apply it before the Cusum limit H is exceeded or is likely to be exceeded by the next observation. In addition to checking for an exceedance of the single-observation maximum during weekly monitoring, use the weekly data to continue the Cusum, and observe whether the Cusum limit H is exceeded.

Table 3.2a: Exceedances Probabilities for Given Numbers of Samples

Probability (for independent, uncorrelated data) of observing data exceeding L at least X times when the true probability is P that data will exceed L. In routine or accelerated monitoring, some level L is set so that L should be exceeded only rarely (P = 0.95 to 0.99). Thus, only rarely (with low probability) will there be one or more observations out of N for which L is exceeded. If the true probability P that data may exceed L is larger, there is a higher probability of observing X = 1, 2, or 3 exceedances of L when taking N observations.

P	N = 3			N = 4			N = 5			N = 6		
	$X \geq 1$	$X \geq 2$	$X \geq 3$	$X \geq 1$	$X \geq 2$	$X \geq 3$	$X \geq 1$	$X \geq 2$	$X \geq 3$	$X \geq 1$	$X \geq 2$	$X \geq 3$
0.10	1.00	0.97	0.73	1.00	1.00	0.95	1.00	1.00	0.99	1.00	1.00	1.00
0.20	0.99	0.90	0.51	1.00	0.97	0.82	1.00	0.99	0.94	1.00	1.00	0.98
0.30	0.97	0.78	0.34	0.99	0.92	0.65	1.00	0.97	0.84	1.00	0.99	0.93
0.40	0.94	0.65	0.22	0.97	0.82	0.48	0.99	0.91	0.68	1.00	0.96	0.82
0.50	0.88	0.50	0.12	0.94	0.69	0.31	0.97	0.81	0.50	0.98	0.89	0.66
0.60	0.78	0.35	0.06	0.87	0.52	0.18	0.92	0.66	0.32	0.95	0.77	0.46
0.70	0.66	0.22	0.03	0.76	0.35	0.08	0.83	0.47	0.16	0.88	0.58	0.26
0.80	0.49	0.10	0.01	0.59	0.18	0.03	0.67	0.26	0.06	0.74	0.34	0.10
0.90	0.27	0.03	0.00	0.34	0.05	0.00	0.41	0.08	0.01	0.47	0.11	0.02
0.95	0.14	0.01	0.00	0.19	0.01	0.00	0.23	0.02	0.00	0.26	0.03	0.00
0.975	0.07	0.00	0.00	0.10	0.00	0.00	0.12	0.01	0.00	0.14	0.01	0.00
0.99	0.03	0.00	0.00	0.04	0.00	0.00	0.05	0.00	0.00	0.06	0.00	0.00

Tabled probabilities were calculated from binomial cumulative probabilities, as 1 - Pr(X-1; N, P).
In selecting a rule ("X in N") for accelerated monitoring, the goal is to balance the chance of a false positive against that of a false negative. Because limits or warning levels are intended to represent 95th to 99th percentiles, the chance of a false positive is given in the rows for P = 0.95, 0.975, and 0.99. The chance of a correct confirmation, when P < .95 is given in the other rows; thus the probability of a false negative when P is truly 0.7 (which means that exceedances are expected to occur 30% of the time) is 0.22 for the rule (N=3, X=2).

3.3 Statistical Characterization of Coal Mine Discharge Loadings

Permit data submitted for EPA's use in development of a coal remining database (EPA, 1999) were used to characterize variability of eastern coal mine discharges. This collection of permit data is not a random or statistically designed sample. The survey solicited data for mine sites that would provide examples of BMP performance. Unpublished reports (Griffiths, in preparation) and published sources (Brady et al., 1998, Hornberger et al., 1990) were also used.

Discharge flows, concentrations, and loadings vary remarkably among monthly or weekly samples, over the course of 1-4 years (Brady et al., 1998, Griffiths, in preparation). Sample coefficients of variation (CV) for iron loadings range approximately from 0.25 to 4.0. Sample CVs for manganese loadings ranged from 0.24 to 3.8. These sample statistics were calculated for a selection of ten mine sites having 42 discharges. The CV for the least variable discharge at each site ranged from 0.24 to 1.5 across the ten sites; the CV for the most variable discharge at each site ranged from 0.85 to 4.7.

There is significant, positive autocorrelation of flows and concentrations in mine discharges over periods of 1-4 weeks (Griffiths, in preparation). From Griffiths' reports, the 4-week autocorrelation of flow is 0.4-0.8. Griffiths did not report statistics for loadings.

Loadings of iron, manganese, and sulfate appear to be approximately distributed lognormally. Normality of log (loading) is often rejected by the Shapiro-Wilk test for the larger samples of 80 to 100 data results, but skewness and kurtosis are reduced considerably by log transformation. For log transformed loadings, the sign of the kurtosis estimates is not consistently positive or negative; about 54 percent of samples have skewness estimates less than one in absolute value, and about 83 percent, less than two. For log transformed loadings, the skewness estimate is more often negative than positive, and about 80 percent of samples have skewness less than one in absolute value. For untransformed loadings, non-normality is often confirmed in small samples and the skewness and kurtosis estimates are usually large and positive.

More important than normality of individual data for parametric hypothesis testing is homogeneity of variances and approximate normality of averages. Log transformation eliminates the relation between variance and mean that is apparent in untransformed loadings data, and so provides a variance stabilizing transformation. T-statistics, calculated for a test of the difference between baseline and post-baseline periods, are normally distributed in aggregate, indicating that the standardized difference in averages is very nearly normal. However, data are still asymmetric after transformation, so suitability of the t-test needs to be evaluated for small samples (e.g., n = 12).

These observations indicate that the usual parametric hypothesis tests can be used to compare the average loadings during and after baseline, if one accounts for autocorrelation.

References

Brady, K.B.C., M.W. Smith, and J. Schueck (editors), 1998. Coal Mine Drainage Prediction and Pollution Prevention in Pennsylvania. Pennsylvania Department of Environmental Protection, publ. no. 5600-BK-DEP2256, October 1998.

Griffiths, J.C., R.J. Hornberger, and M.W. Smith (in preparation). Statistical Analysis of Abandoned Mine Drainage in the Establishment of the Baseline Pollution Load for Coal Remining Permits. U.S. Environmental Protection Agency, Washington, D.C.

Hornberger, R.J. et al., 1990. Acid Mine Drainage from Active and Abandoned Coal Mines in Pennsylvania. Chapter 32 of Water Resources in Pennsylvania: Availability, Quality, and Management, Pennsylvania Academy of Science, pp. 432-451.

Loftis, J.C. and R.C.Ward 1980. Sampling Frequency Selection for Regulatory Water Quality Monitoring, Water Resources Bulletin, Vol.16, No. 3, pp.501-507.

U.S. EPA, 1993. Statistical Support Document for Proposed Effluent Limitations Guidelines and Standards for the Pulp, Paper, and Paperboard Point Source Category. EPA publ. no. 821/R-93-023.

U.S. EPA, 1999. Office of Water. Coal Remining Database: 61 State Data Packages, March 1999. Details provided by the U.S. Environmental Protection Agency's Sample Control Center, operated by DynCorp I&ET, 6101 Stevenson Avenue, Alexandria, VA 22304.